

Supervisory Stress Test Models Glossary

The below entries are generally drawn or adapted from selected textbooks and references in econometrics and statistics.¹

- *Autocorrelation*: In a time series or panel data model, correlation between the errors in different time periods.
- *Autoregressive process*: A time series model whose current value depends linearly on one or more of its previous values plus an unpredictable disturbance.
- *Bayesian Information Criterion (BIC)*: A statistic used to measure a model's goodness of fit while penalizing complexity. BIC will favor increased complexity only if the resulting improvement in fit sufficiently compensates for the loss of simplicity.
- *Binary random variable*: A random variable that takes on the values zero or one.
- *Collinearity*: A term that refers to correlation among the independent variables in a multiple regression model; it is usually invoked when some correlations are "large," but an actual magnitude is not well defined.
- *Correlation*: More formally, "linear correlation": the linear dependence between two random variables.
- *Covariate*: See *independent variable*.
- *Cumulative distribution function*: A function that gives the probability of a random variable being less than or equal to any specified real number.
- *Degrees of freedom*: In multiple regression analysis, the number of observations minus the number of estimated parameters.
- *Dependent variable*: The variable to be explained in a model (such as a linear regression model).

¹ *Introductory Econometrics: A Modern Approach* by J. Wooldridge; *Microeconometrics: Methods and Applications* by A.C. Cameron and P.K. Trivedi; *The Cambridge Dictionary of Statistics* by B.S. Everitt and A. Skrondal; "Nonparametric Estimation," Wolfram MathWorld, Wolfram, August 25, 2025, <https://mathworld.wolfram.com/NonparametricEstimation.html>

- *Endogeneity*: The presence of an explanatory variable in a multiple regression model that is correlated with the error term, either because of an omitted variable, measurement error, or simultaneity.
- *Error term*: The variable in a simple or multiple regression equation that contains unobserved factors which affect the dependent variable. The error term may also include measurement errors in the observed dependent or independent variables.
- *Generalized method of moments (GMM)*: A flexible method of estimating models that extends upon simpler methods such as ordinary least squares regression and thus entails fewer assumptions about the data than these simpler methods. GMM allows estimated relationships to be linear or nonlinear.
- *Indicator variable*: A variable that takes on the value zero or one depending on whether a certain condition is met.
- *Heteroskedasticity*: Describes situations where the variance of the error term, given the explanatory variables, is not constant. If not accounted for, heteroskedasticity can make certain common statistical tests invalid.
- *Independent variable*: In regression analysis, a variable that is used to explain variation in a dependent variable.
- *Intercept parameter*: The parameter in a multiple regression model that gives the expected value of the dependent variable when all the independent variables equal zero.
- *Latent variable model*: A model where the observed dependent variable is assumed to be a function of an underlying latent, or unobserved, variable.
- *Linear interpolation*: The process of determining the value of a function between two known values by assuming a linear relationship between them.
- *Linear regression model*: A model linear in its parameters, where the dependent variable is a function of independent variables plus an error term.
- *Logarithmic function*: A mathematical function defined for positive arguments that has a positive, but diminishing, slope.
- *Logit model*: A model for a binary dependent variable where the probability of the outcome is modeled using a non-linear function of the independent variables; specifically, an “S”-shaped curve known as the logistic function.
- *Maximum likelihood estimator*: A broadly applicable estimation method where the parameter estimates are chosen to maximize the log-likelihood function.

- *Monotonicity*: A term describing a sequence of values that never decreases (“monotonic increasing”) or that never increases (“monotonic decreasing”).
- *Negative binomial distribution*: The probability distribution of the number of failures before a specified number of successes in a sequence of draws of a binary random variable, where “success” and “failure” are defined as the binary random variable taking the values of 1 or 0.
- *Nonparametric model*: A model that allows the functional form of a fit to data to be obtained in the absence of any guidance or constraints from theory. This class of models is extremely flexible but does not produce parameter estimates that summarize relationships between variables.
- *Null hypothesis*: In classical statistical testing, a hypothesis is taken to be true unless the data provides substantial evidence against it. Typically, this is a hypothesis that there is no relationship between two variables.
- *Ordinary least squares regression*: A method for estimating the parameters of a multiple linear regression model. The ordinary least squares estimates are obtained by minimizing the sum of squared residuals.
- *Overfit*: A term describing a model that contains more unknown parameters than can be justified by the data.
- *P-value*: The smallest significance level at which the null hypothesis can be rejected. Equivalently, the largest significance level at which the null hypothesis cannot be rejected.
- *Panel data*: A data set constructed from repeated cross sections over time.
- *Panel regression model*: A regression model estimated on panel data.
- *Parsimonious model*: A model with as few parameters as possible for capturing any desired features.
- *Pseudo R-squared*: A goodness-of-fit measure for models of binary random variables.
- *R-squared*: In a multiple regression model, the proportion of the total sample variation in the dependent variable that is explained by the independent variables.
- *Random variable*: A variable whose outcome is uncertain.
- *Regression coefficient*: In a linear regression model, the change in the dependent variable when the associated independent variable increases by one. In other forms of regression, such as logit models, coefficients may have more complex interpretations, but still serve to summarize the relationship between the independent and dependent variables.

- *Residual*: The difference between the actual and the fitted (or predicted) value from a regression model.
- *Root mean squared error*: In multiple regression analysis, the estimate of the standard deviation of the population error, obtained as the square root of the sum of squared residuals over the degrees of freedom.
- *Standard error*: Generically, an estimate of the standard deviation of an estimator.
- *Statistical significance*: Rejecting the null hypothesis that a parameter is equal to zero against the specified alternative, at the chosen significance level.
- *Taylor series approximation*: The approximation of a function near a specific point as the sum of a simpler function and a remainder. This simpler function is from a class of functions known as “polynomials,” such as a quadratic or cubic function.
- *Tobit model*: A model for a dependent variable that takes on the value zero with positive probability but is roughly continuously distributed over strictly positive values.
- *Unconditional distribution*: The probability distribution of a random variable, defined without reference to the values of any other random variables.
- *Vector*: A quantity with both magnitude and direction.
- *Vector Autoregressive (VAR) Model*: A model for two or more time series where each variable is modeled as a linear function of past values of all variables, plus disturbances that have zero means given all past values of the observed variables.
- *Weighted least squares estimator*: A type of linear regression estimator used to adjust for a known form of heteroskedasticity, where each squared residual is weighted by the inverse of the (estimated) variance of the error.
- *Winsorization*: A procedure that moderates the influence of outlier points in a data sample to increase the robustness of model parameter estimates. The procedure consists of replacing datapoints above (or below) a certain value, usually defined as a percentile of the sample distribution. For example, a 5% Winsorization would replace values above the 95th percentile, or below the 5th percentile, with the values at those percentiles of the sample data.